EXHIBIT C

Protein microarrays and proteomics

Gavin MacBeath

doi:10.1038/ng1037

The system-wide study of proteins presents an exciting challenge in this information-rich age of whole-genome biology. Although traditional investigations have yielded abundant information about individual proteins, they have been less successful at providing us with an integrated understanding of biological systems. The promise of proteomics is that, by studying many components simultaneously, we will learn how proteins interact with each other, as well as with non-proteinaceous molecules, to control complex processes in cells, tissues and even whole organisms. Here, I discuss the role of microarray technology in this burgeoning area.

As the scope of proteomics is extremely broad, it is helpful to distinguish among its various aspects before discussing how microarray technology is likely to feature. Over the past decade, the marriage of two technologies—two-dimensional gel electrophoresis and mass spectrometry—has spawned what I refer to as 'unbiased' or 'discovery-oriented proteomics' (Fig. 1). In a typical discovery-oriented experiment, a biological sample is analyzed by separating, quantifying and identifying as many proteins as possible, often with an emphasis on those proteins with altered abundance relative to a reference sample. The approach is unbiased because the investigator does not impose his or her knowledge of biology on the experimental design; it is discovery-oriented because unknown proteins, as well as known proteins, may be identified in the course of the experiment.

Between this and traditional biochemistry lies 'focused' or 'systems-oriented proteomics' (Fig. 1). Here, the investigator directs the study by defining a subset of proteins to be analyzed, such as a family of proteins related by sequence or a collection of proteins related by function. Although microarray technology will undoubtedly have an important role in discovery-oriented proteomics, it is particularly well suited to systems-oriented investigations.

Another important distinction in proteomics concerns the objective of the study. In one type of experiment, the investigator aims to analyze a particular biological sample such as a tumor biopsy from an individual with cancer. For such experiments, a complete 'proteomic analysis' would involve measuring the abundance, modification, activity, localization and interaction of all the proteins in that sample. In practice, current technologies limit our analysis to only one or two of these parameters and to only a fraction of the proteins. This type of experiment is the protein equivalent of expression profiling and is broadly known as 'protein profiling'. I refer to this aspect of proteomics as 'quantitative proteomics' to distinguish it from the second type of investigation, 'functional proteomics', in which one seeks instead to define the function of every protein in a given organism. This may involve, for example, identifying putative substrates for enzymes or putative interactions between proteins. It is already clear that microarray technology will have a significant impact on both quantitative proteomics and functional proteomics, and in this review I address each of these areas in turn.

Quantitative proteomics

Microarray technology is finding its way into quantitative proteomics through the construction of what are most accurately called 'protein-detecting microarrays'[1]. A protein-detecting microarray comprises many different affinity reagents (frequently antibodies) arrayed at high spatial density on a solid support. Each agent captures its target protein from a complex mixture (such as serum or cell lysate), and the captured proteins are subsequently detected and quantified.

Historically, the roots of protein-detecting microarrays lie in the development of immunoassays. As early as 1929, antibodies were used in serology to precipitate antigens for subsequent quantification[2]. Analytical immunoassay technology was greatly advanced with the introduction of the radioimmunoassay (RIA) by Yarlow and Berson[3] in 1959 and the enzyme-linked immunosorbant assay (ELISA) by Engvall and Perlman[4] in 1971. In the late 1980s, Roger Ekins argued that extremely sensitive, quantitative assays could, in principle, be developed using 'microspots' of antibodies on solid supports and that multianalyte assays could be carried out using spatially separated arrays of such spots[5]. But it was not until the late 1990s—when the benefits of DNA microarray technology were well recognized and the equipment to manufacture microarrays became more accessible—that attention shifted to the development of antibody microarrays for the purpose of quantitative proteomics.

Sandwich immunoassays. One of the earliest reports demonstrating the feasibility of multianalyte, microarray-based immunoassays was published by Silzel *et al.*[6] in 1998. Using a standard inkjet printer, they spotted monoclonal antibodies directed against the four human immunoglobulin-γ (IgG) subclasses onto a thin sheet of polystyrene film to form spots with a diameter of 200 μm. Using a sandwich-based assay (Fig. 2*a*), they were able to show subclass-specific recognition of human myeloma proteins with minimal crossreactivity and to observe dose-dependent signals for each subclass. Although the spotting methods were relatively crude and difficult to scale up, this early study showed that several proteins could be detected and quantified in a parallel fashion using antibody microarrays.

In one of the first studies to show that multiplexed sandwich immunoassays could be used to analyze biological samples, Moody *et al.*[7] fabricated arrays of seven different antibodies

Department of Chemistry and Chemical Biology, and Bauer Center for Genomics Research, Harvard University, 12 Oxford Street, Cambridge, Massachusetts 02138, USA (e-mail: macbeath@chemistry.harvard.edu).

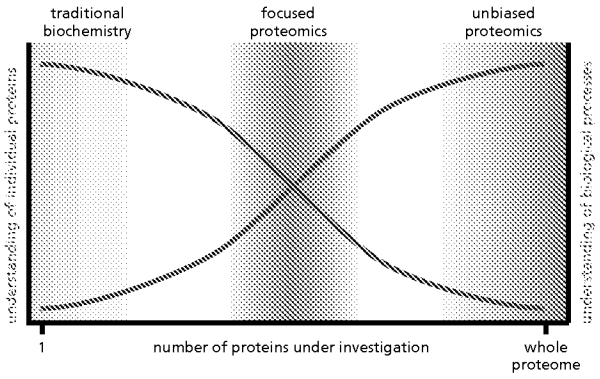

Fig. 1 The spectrum of protein analysis, ranging from traditional biochemistry to unbiased proteomics. Focused proteomics occupies a middle ground, where one seeks to maximize both the depth and the breadth of biochemical investigation.

against cytokines in a 3 × 3 pattern in 96-well polystyrene microtiter plates. They then used these arrays to monitor the abundance of cytokine production by THP-1 cells on stimulation with lipopolysaccharide, either alone or in combination with the anti-inflammatory drug dexamethasone. Captured cytokines were probed with a cocktail of detection antibodies, and the signal was amplified by enzyme-catalyzed chemiluminescence. This provided the sensitivity needed to detect cytokine concentrations of less than a picogram per milliliter. Shortly thereafter, Huang *et al.*[8] reported fabricating hand-spotted arrays of antibodies on nitrocellulose membranes and showed that they could detect up to 24 different cytokines in both conditioned media and human serum in a multiplexed sandwich assay, also using a chemiluminescent detection system. Several other groups have subsequently reported antibody arrays that measure human cytokines (see, for example, refs 9–11).

In the most extensive study carried out so far, Schweitzer *et al.*[12] fabricated microarrays of 75 antibodies against cytokines on chemically derivatized glass slides and used these arrays to study cytokine secretion from human dendritic cells induced with lipopolysaccharide or tumor-necrosis factor-α. To increase the sensitivity of their assay and retain high spatial resolution (not afforded by enzyme-catalyzed chemiluminescence), they used isothermal rolling-circle amplification[13] to develop their sandwich assays. Particularly notable is the extent to which they were able to multiplex this technique. Some of the detection antibodies were found to crossreact with other cytokines on the array. To avoid this problem, Schweitzer *et al.*[12] divided their collection of capture antibodies into two roughly equal-sized groups. Each group of antibodies was spotted on a different portion of the same glass slide separated by a Teflon barrier. In this way, spurious signals generated by antibody crossreactivity could be avoided. This strategy of subdivision represents a potentially general solution to some of the problems inherent in multiplexed sandwich assays (see below).

It is no coincidence that almost every sandwich microarray-based assay that has been reported is directed at the analysis of human cytokines. There are two reasons for this. First, matched pairs of high-affinity, high-specificity antibodies have been developed and commercialized over the past several decades for use in diagnostic kits based on sandwich ELISAs. The availability of high-performance, validated reagents circumvents the largest hurdle in developing microarray-based assays. Second, cytokines are secreted from cells; thus, analyses are done on relatively simple mixtures in comparison to cellular lysates.

In experiments carried out in collaboration with Peter Sorger, my research group has found that it is much more difficult to find antibodies to study intracellular proteins. In fact, only about 5% of over 100 commercial antibodies that we have tested are suitable for microarray-based analyses of cellular lysates (U.B. Nielsen, M.H. Cardone, A.J. Sinskey, G.M. and P.K. Sorger, unpublished data). Nevertheless, we have been able to develop arrays that report on both the abundance and the modification state of intracellular signaling proteins, including integral membrane receptors. We have also developed a ratiometric quantification method in which one detection antibody, labeled with a fluorophore, reports on the 'abundance' of the captured protein, while a second detection antibody, labeled with a different colored fluorophore, reports on a specific post-translational 'modification' state of that protein. The resulting ratio of fluorescence provides a measure of the fraction of protein in the lysate that has been modified (for example, phosphorylated; U.B. Nielsen, M.H. Cardone, A.J. Sinskey, G.M. and P.K. Sorger, unpublished data). We anticipate that this and similar strategies will be particularly useful for analyzing complex signaling networks on a system-wide basis.

Antigen capture immunoassays. In an alternative approach to the sandwich assay, it is possible to detect captured antigens without using a second antibody (Fig. 2*b*). This requires either a label-free detection method (such as mass spectrometry or surface plasmon resonance) or the chemical labeling of all proteins in a sample. In one of the earliest and most extensive attempts to examine the feasibility of a two-color labeling approach, Haab *et al.*[14] assembled a collection of 115 antibodies and their protein ligands. They prepared defined mixtures of the antigens, labeled the mixtures with activated Cy5 dye, and combined each mixture with a Cy3-labeled reference mixture. The proteins were then applied to antibody microarrays comprising the 115 antibodies spotted at high density onto slides coated with poly-L-lysine. Haab *et al.*[14] found that only 20% of the arrayed antibodies provided specific and accurate measurements of their target antigens at a concentration of 1.6 μg/ml or less, but that some antibodies could detect ligands at concentrations of less than 1 ng/ml and at partial concentrations of 1 part in 1,000,000.

This ratiometric, two color approach was recently applied to the proteomic analysis of LoVo colon carcinoma cells in response to treatment with ionizing radiation[15]. Proteins in lysates from untreated cells and cells collected 4 h after irradiation were labeled with Cy3 and Cy5, respectively. The samples were mixed and applied to a microarray composed of 146 distinct antibodies directed against proteins involved in stress response, cell-cycle progression and apoptosis. In addition to observing the upregulation of five apoptotic proteins already known to be induced by radiation, Sreekumar *et al.*[15] discovered six other proteins that were upregulated and one that was downregulated; most of these results were confirmed subsequently by immunoblotting.

This study is significant because it shows that protein-detecting microarrays can provide new information about previously characterized proteins and thus can be used to generate fresh hypotheses. But it is important to note that, although many different antibodies were arrayed, the researchers did little to validate their performance in a microarray format. It is well known that most antibodies crossreact significantly with other cellular proteins, but none of the antibodies were tested for crossreactivity using cellular lysates. These arrays must therefore be seen as tools for discovery rather than as tools for quantitative analysis. One must also view the number of elements on their arrays with some skepticism: although 146 antibodies were immobilized, it is not clear what percentage of these antibodies generated meaningful data.

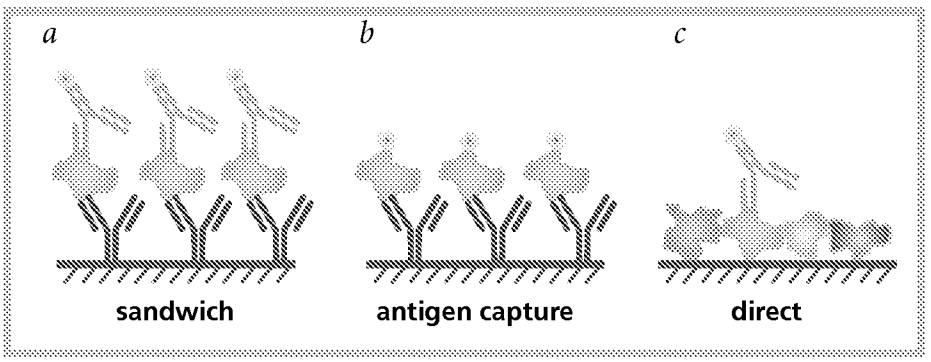

Fig. 2 Current immunoassay strategies used in protein-detecting microarrays. ***a***, In a sandwich immunoassay, capture antibodies are immobilized on the solid support, and bound proteins are detected using a second, labeled detection antibody. ***b***, In an antigen capture assay, proteins are similarly captured by immobilized antibodies, but the captured proteins are detected directly. This is usually accomplished by chemically labeling the complex mixture of proteins before applying them to the array. In the two-color version of this assay, two samples are labeled independently with distinguishable fluorophores, and the samples are mixed before applying them to the array. **c**, In a direct assay, the complex mixture of proteins is itself immobilized on the solid support, and specific proteins in that mixture are visualized using labeled detection antibodies.

In principle, the two-color approach permits more accurate quantification of the relative abundance of proteins, but it is hampered by an inherent lack of sensitivity. Because the approach does not offer signal amplification beyond that provided by the use of fluorophores, the limit of detection is typically no better than 1 ng/ml of analyte. Enzyme-catalyzed signal amplification improves the limit of detection by at least three orders of magnitude. Taking advantage of this increased sensitivity at the cost of adopting a one-color approach, Knezevic *et al.*[16] have used microarrays of 368 distinct antibodies, coupled with laser capture microdissection, to profile proteins in both normal and cancerous epithelium in the oral cavity, as well as in stromal cells adjacent to and surrounding these tissues. Proteins in cellular lysates were biotinylated and applied to the arrays, and the captured proteins were detected using an enzyme-linked colorimetric assay. As with the study of Sreekumar *et al.*[15], these arrays must be seen as tools for discovery rather than quantitative analysis because little attempt was made to evaluate the performance of each antibody. Although only 14% of the antibodies on the arrays gave consistently positive signals with any of the oral cavity tissues, Krizman and co-workers were nevertheless able to identify 11 proteins that reproducibly changed in either their relative abundance or their state of phosphorylation with respect to disease progression. The discovery of multiple markers of disease progression and their facile analysis using protein-detecting microarrays holds great promise for diagnosis, prognosis and tailored medicine[17].

In an interesting twist on the issue of direct detection, Belov *et al.*[18] have used a microarray of 60 antibodies directed against cluster of differentiation (CD) antigens to immunophenotype leukocytes obtained from both normal individuals and individuals with various types of leukemia. Because these antigens are expressed on the surface of cells, the researchers could apply whole cells to their arrays and detect captured cells using dark-field microscopy. From the expression pattern of the CD antigens, they defined a fingerprint that is diagnostic of chronic lymphocytic leukemia and are currently using their array to establish fingerprints for other types of leukemia.

Direct immunoassays. In a third approach to protein profiling, it is possible to immobilize the samples themselves, rather than the affinity reagents, on the solid support. Specific proteins in these samples can then be detected and quantified by probing with labeled antibodies. Paweletz *et al.*[19] have used this approach to analyze pro-survival checkpoint proteins in patient-matched normal prostate epithelium, prostate intraepithelial neoplasia and invasive prostate cancer. All three types of cell were isolated from 10 individuals by laser capture microdissection. The cells were lysed, arrayed on nitrocellulose-coated slides and probed with different antibodies. Paweletz *et al.*[19] found that the abundance of phosphorylated Erk decreased during disease progression, whereas the abundance of phosphorylated Akt tended to increase. Control experiments suggested that the lower limit of detection in their assay was around 1 ng/ml; however, because these experiments were done with purified antigen rather than with complex lysates, the actual limit of detection is likely to be considerably higher. Using a similar approach, Madoz-Gurpide *et al.*[20] have described the benefits of simplifying cell lysates chromatographically before arraying them on a solid support. This strategy increases the specificity of direct immunoassays, but at the cost of substantially decreasing throughput.

In a different vein, microarrays of defined antigens represent a special type of direct immunoassay that can be used to profile complex mixtures containing immunoglobulins. For these experiments, the immobilized antigens direct the 'capture' of solution-phase immunoglobulins, although in reality it is the immunoglobulins that are recognizing the antigens. Early experiments by Mendoza *et al.*[21] showed that protein antigens spotted at very high density onto chemically activated glass surfaces could be detected specifically with solution-phase antibodies. More recently, microarrays of either purified antigens or antigen-containing extracts have been used to profile allergen-specific IgEs from human serum[22–26], to characterize and quantify autoantibody responses in individuals with autoimmune diseases[27,28], and to serodiagnose people with parasitic or viral infections[29].

Advantages and limitations of protein-detecting microarrays. In contrast to relatively unbiased separation methods such as gel electrophoresis and liquid chromatography, affinity-based approaches enable the investigator to direct the experiment. This is why microarray technology is best suited to systems-oriented investigations. If the goal of the experiment is to study a particular biological process, only those proteins involved in that process need to be examined. The downside, of course, is that these investigations require both a prior knowledge of the proteins to be studied and appropriate affinity reagents. This highlights the importance of efforts directed at the rapid selection of recombinant antibodies, antibody substitutes and small-molecule ligands, and it will be interesting to see how these technologies develop over the next few years.What, then, are the relative merits of each type of protein-detecting microarray, and which approach is best? In my (current) estimation, direct immunoassays are not likely to see widespread use. Although this format readily permits signal amplification, proteins in low abundance are difficult to observe because they must compete with more abundant proteins for immobilization on the solid support. The technique, like the dot-blot that it mimics, is also

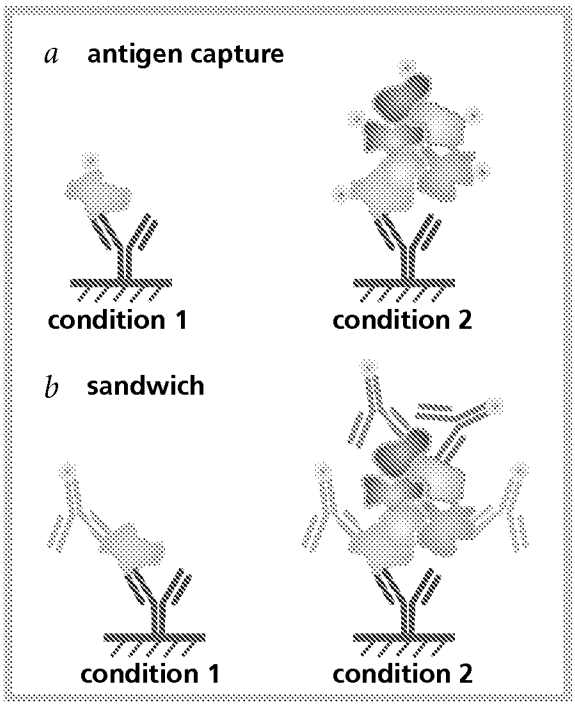

Fig. 3 Complications arising from regulated protein–protein interactions. ***a***, In an antigen capture assay, a more intense signal is observed when the target protein is involved in a multiprotein complex (condition 2) than when it is free in solution (condition 1). ***b***, In a sandwich assay, a similar problem occurs if detection antibodies are present that recognize other components of the multiprotein complex.

limited by the specificity of the affinity reagents (see below), and the accuracy of quantification is limited by the reproducibility of sample preparation and spotting.

Two-color antigen capture methods are less demanding in terms of sample preparation and spotting because quantification involves a ratiometric readout. In its current form, the two-color method does not involve signal amplification (beyond that afforded by the use of fluorescent dyes) and thus has limited sensitivity (~1 ng/ml). This may be circumvented, however, by the use of two orthogonal enzyme-catalyzed amplification methods or by two-color rolling-circle amplification. Like the direct assay, this approach is limited by the specificity of the affinity reagents. In addition, the need to label proteins presents a confounding problem: some proteins may be labeled preferentially on their antigenic epitopes and so lose their ability to be captured by their affinity reagents. For example, my colleagues and I have encountered proteins that are not detected using antigen capture methods but are readily detected using the same capture antibodies in a sandwich-based assay (U.B. Nielsen, M.H. Cardone, A.J. Sinskey, G.M. and P.K. Sorger, unpublished data).

Unlike the antigen capture method, the sandwich approach does not require the proteins to be labeled. This simplifies sample preparation and vastly increases throughput, which is why this strategy is likely to dominate array-based diagnostics. In addition, the sandwich approach has the distinct advantage that it addresses the largest problem confronting protein-detecting microarrays—that is, specificity. As anyone who has carried out an immunoblot knows, most antibodies crossreact with other cellular proteins. Both the direct assay approach and the antigen capture approach are analogous to a western blot in which each lane has been compressed into a single band; all the crossreacting proteins contribute to the total signal. By contrast, a sandwich assay is analogous to an immunoprecipitation reaction coupled with a compressed western blot. One is far more likely to observe a single band on a western blot when the product of an immunoprecipitation reaction is analyzed than when total cellular lysate is analyzed. The disadvantage of the sandwich approach, however, is that two noncompeting affinity reagents are required for each protein.

Complications from regulated protein–protein interactions. A potential caveat of protein-detecting microarrays is that regulated protein–protein interactions complicate the interpretation of data arising from either antigen capture or sandwich assays. For example, a protein may be free in solution under one set of conditions and form part of a large, multiprotein complex under another set. Although the abundance of the protein does not actually change in going from the first set of conditions to the second, an antigen capture assay would report a significant increase (Fig. 3*a*)[1]. Even if the individual carrying out the experiment were aware of this interaction, it is not clear how she or he would distinguish upregulation from complex formation.

In a multiplexed sandwich assay, two potential problems confront the investigator. First, the epitope recognized by the detection reagent may become obscured in the complex, resulting in an apparent decrease in signal on complex formation. Second, the presence of detection antibodies directed against other members of the complex would result in an apparent increase in signal (Fig. 3*b*). In either case, however, knowledge of the biology would enable the investigator to circumvent these problems. Detection antibodies could be chosen that do not target protein–protein interfaces, and arrays could be subdivided physically to avoid the detection of associated proteins. These complications highlight the need to understand the underlying function of the proteins under investigation.

Functional proteomics

Genome sequencing projects contribute daily to lists of the protein components of cells and organisms; determining what these components do is the task of functional proteomics. As part of this endeavor, considerable effort is now being devoted to defining protein–protein interactions on a system-wide or genome-wide basis. One technique that has been used extensively is the yeast two-hybrid system[30–32]. This *in vivo* assay, although easy to implement and of considerable utility, has several limitations. Proteins that function as transcriptional activators yield false positives when fused to DNA-binding domains, whereas false negatives arise when proteins are displayed inappropriately or fail to fold correctly in yeast. More significantly, the investigator is unable to control either the post-translational modification state of the proteins under investigation or the environment under which the interactions are being studied.

Recently, immunoprecipitation coupled with mass spectrometry has been used to identify multiprotein complexes on a large scale in *Saccharomyces cerevisiae*[33,34]. This technique, which can be adapted readily to high-throughput investigations, identifies proteins that purify together with an epitope-tagged bait protein. Although this is a very efficient way to identify protein complexes, it has the caveat that a single bait protein may occur in more than one complex in a cell. It may therefore bring down two or more proteins that never actually colocalize, giving the illusion that protein complexes are bigger and more elaborate than they really are.

As a complement to these approaches, protein microarray technology provides a well-controlled, *in vitro* way to study function on a system-wide or genome-wide basis. For this application, the proteins themselves, rather than affinity reagents, are arrayed on a solid support. To distinguish these types of arrays

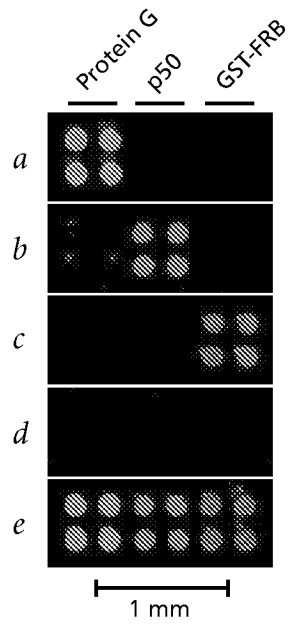

Fig. 4 Proof-of-concept experiments showing the detection of protein–protein interactions on a model protein function microarray[37]. Three different proteins were spotted in quadruplicate on each of five glass microscope slides. ***a***, Slide probed with IgG labeled with BODIPY-FL (a fluorophore); blue spots indicate an interaction of IgG with immobilized protein G. ***b***, Slide probed with Cy3-labeled IκBα; green spots indicate an interaction of IκBα with immobilized p50. ***c***,***d***, Slides probed with Cy5-labeled FKBP12 in the presence (*c*) or absence (*d*) of 100 nM rapamycin; red spots indicate the rapamycin-dependent interaction of FKBP12 with immobilized FRB. ***e***, Slide probed with a mixture of all three labeled proteins and 100 nM rapamycin.

from protein-detecting microarrays, I refer to them (using Kodadek's nomenclature[1]) as 'protein function microarrays'.

An advantage of studying proteins in an array format is that the investigator can control the conditions of the experiment. This includes not only factors such as pH, temperature, ionic strength and the presence or absence of cofactors, but also the modification states of the proteins under investigation. An early study by Ge[35] illustrates this principle. Ge prepared low-density arrays composed of 48 purified human proteins spotted on nitrocellulose membranes and probed his arrays with several different radiolabeled molecules including proteins, nucleic acids and small organic compounds. Because his arrays featured both a phosphorylated and nonphosphorylated form of the protein PC4, he was able to show that a double-stranded DNA probe was bound more tightly by the phosphorylated form of the protein than by the unmodified one.

Although low-density arrays such as these are useful, there is substantial benefit to be gained from using microarray technology. First, in principle thousands of proteins can be spotted on a single slide or similar support, enabling one to interrogate simultaneously the function of many different proteins with minimal sample consumption. Second, hundreds or even thousands of copies of an array can be fabricated in parallel, enabling the same proteins to be probed repeatedly with many different molecules under many different conditions. It is the combination of these two features that makes microarray technology so well suited to systems-oriented proteomics.

In one of the earliest reports describing methods to fabricate microarrays of functionally active proteins, Arenkov *et al.*[36] spotted purified enzymes onto glass-supported micromatrices of polyacrylamide gel pads and showed that the immobilized proteins retained their catalytic activity. Shortly thereafter, Stuart Schreiber and I reported simplified methods to fabricate and process protein function microarrays using commercially available arrayers and scanners[37]. We showed that purified proteins retain their activities when spotted onto chemically derivatized glass slides featuring amine-reactive, hydrophilic surfaces. Nanoliter volumes of purified proteins were spotted at high spatial densities, permitting over ten thousand spots to fit on a single glass microscope slide. Using surfaces displaying aldehyde groups or featuring a covalently bound layer of chemically activated serum albumin, we showed that protein function microarrays can be used to detect stable interactions between two proteins, transient interactions between enzymes and their substrates, and both high and low affinity interactions between proteins and small molecules (see, for example, Fig. 4). A key feature of this study was that the reported methods could be carried out with standard instrumentation, making protein microarray technology readily accessible for functional proteomics.

Protein function microarrays for focused proteomics. Array technology provides an ideal way to study focused collections of proteins or protein domains. In a study illustrating this concept, Zhu *et al.*[38] cloned, expressed and purified 119 of the 122 predicted protein kinases in yeast with the goal of analyzing their substrate specificities. To reduce sample consumption, they fabricated what were, in essence, miniaturized microtiter plates: 'chips' featuring 10 × 14 wells with a pitch of 1.8 mm. They covalently coated 16 of these chips with 16 different protein or peptide substrates and left a seventeenth chip uncoated to study autophosphorylation. They then manually pipetted a different purified kinase, together with [γ-^{33}P]ATP, into each well. The wells were washed and exposed to a phosphorimager to detect putative substrates. As well as observing different degrees of specificity among the various kinases, Zhu *et al.*[38] discovered that 27 of the 119 kinases could phosphorylate poly(Tyr-Glu).

On the basis of sequence analysis, yeast do not encode any members of the conventional tyrosine kinase family; however, the experiments of Zhu *et al.*[38] suggest that, by analyzing a large set of yeast kinases *in vitro*, it may be possible to predict tyrosine kinases on the basis of their biochemical function. As this is a rather controversial subject, it will be interesting to see how many of these kinases actually phosphorylate tyrosine residues on their *in vivo* substrates.

Protein function microarrays for discovery-oriented proteomics. Protein microarrays can also be used to study function on a genome-wide basis. With this as their goal, Snyder and co-workers[39] cloned 5,800 of the roughly 6,200 yeast open reading frames into a yeast high-copy expression vector carrying an amino-terminal tag of glutathione *S*-transferase (GST) plus oligohistidine. They expressed the proteins in 96-well blocks and purified them by affinity chromatography in microtiter plates. The purified material was then spotted in duplicate onto nickel-coated microscope slides to produce what they termed a yeast 'proteome chip' (Fig. 5). When they probed one of their chips with labeled calmodulin, they were able to identify 6 of the 12 known yeast calmodulin targets and an additional 33 potential targets.

A unique advantage of protein function microarrays is that they can be used to study the interaction of proteins with non-proteinaceous molecules, including nucleic acids, lipids and small organic compounds. To identify yeast proteins that bind closely related phospholipids, Snyder and co-workers[39] probed six of their proteome chips with labeled liposomes, identifying 150 putative phospholipid-binding proteins. By comparing the results from each independent experiment, they were able to determine which interactions were specific for a given lipid and which interactions were not specific. Curiously, of the 48 proteins in yeast that are predicted to contain either PH, PX or FYVE domains (the best-characterized of the lipid-binding domains[40]), only 1 protein (OPY1) was identified in their screens. It is likely that many more yeast proteins recognize the phospholipids under investigation, but were either not present in sufficient amounts on the arrays or not functional. Nevertheless,

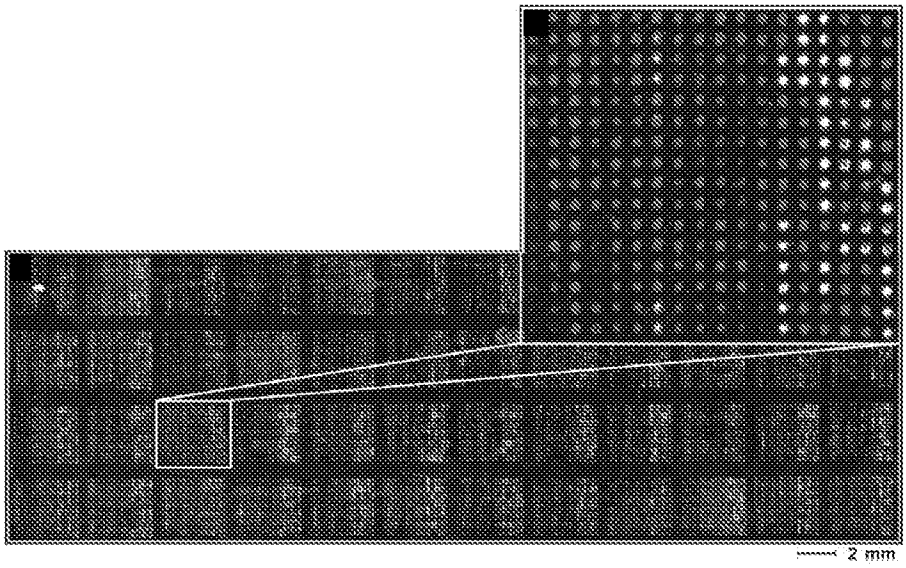

Fig. 5 A 'proteome chip' composed of 6,566 protein samples representing 5,800 unique proteins, which are spotted in duplicate on a single nickel-coated glass microscope slide[39]. The immobilized GST fusion proteins were detected using a labeled antibody against GST.

the ability to assess specificity represents a key feature of protein array technology that will no doubt prove invaluable in future whole-proteome experiments. By the same token, protein arrays should prove useful for evaluating the specificity of lead compounds in drug discovery efforts, for identifying secondary targets of bioactive small molecules, and even for evaluating specificity among a set of related targets in a high-throughput screening format.

Future prospects

Protein function microarrays will benefit undoubtedly from improved methods of fabrication, processing and analysis, but at present the greatest obstacle barring their widespread use is the production of large collections of pure, recombinant proteins. Such production requires, as a necessary first step, the cloning of cDNAs. One approach has been described in which cDNA libraries are screened for protein-producing clones using high-throughput methods[41,42]. Although this is a potentially economical way to generate large collections of proteins and has been used for some discovery-oriented applications[43,44], it is limited in several respects. The resulting libraries are not normalized, they fail to contain genes that are expressed at low abundance, and the proteins often begin and end at random positions in the coding sequence. With the availability of whole-genome sequence data, systematic efforts are now underway to prepare perfectly normalized, indexed collections of full-length open reading frames in recombination-based cloning vectors[45,46]. Resources such as these, coupled with high-throughput methods of expressing and purifying recombinant proteins[47,48], should accelerate greatly the application of microarray technology to functional proteomics.

So what can we expect in the coming years? At the very least we can anticipate continued innovation. Numerous alternative types of array have been reported already, most notably bead-based 'virtual' arrays[49] and live, transfected cell arrays[50]. Each approach brings with it both advantages and limitations. Regardless of the approach, there is a growing need in array-based proteomics for an increased emphasis on quality rather than on numbers. With respect to protein-detecting microarrays, the crossreactivity of affinity reagents needs to be assessed and reported. For protein function microarrays, the purity and integrity of the proteins need to be determined, and their concentrations need to be normalized. Perhaps the biggest challenge is to find a feasible way to assess the functional state of purified proteins. Solubility alone, although informative, does not indicate that a protein is folded correctly or processed appropriately. By paying close attention to these issues, data generated by both systems-oriented analyses and discovery-oriented expeditions should lead us ultimately toward a more integrated understanding of biological processes.

1. Kodadek, T. Protein microarrays: prospects and problems. *Chem. Biol.* **8**, 105–115 (2001).
2. Heidelberger, M. & Kendall, F.E. A quantitative study of the precipitation reaction between type III pneumococcus polysaccaride and purified homologous antibody. *J. Exp. Med.* **50**, 809–823 (1929).
3. Yalow, R.S. & Berson, S.A. Assay of plasma insulin in human subjects by immunological methods. *Nature* **184**, 1648–1649 (1959).
4. Engvall, E. & Perlman, P. Enzyme-linked immunosorbent assay (ELISA). Quantitative assay of immunoglobulin G. *Immunochemistry* **8**, 871–874 (1971).
5. Ekins, R., Chu, F. & Biggart, E. Multispot, multianalyte, immunoassay. *Ann. Biol. Clin. (Paris)* **48**, 655–666 (1990).
6. Silzel, J.W., Cercek, B., Dodson, C., Tsay, T. & Obremski, R.J. Mass-sensing, multianalyte microarray immunoassay with imaging detection. *Clin. Chem.* **44**, 2036–2043 (1998).
7. Moody, M.D., Van Arsdell, S.W., Murphy, K.P., Orencole, S.F. & Burns, C. Array-based ELISAs for high-throughput analysis of human cytokines. *Biotechniques* **31**, 186–194 (2001).
8. Huang, R.P., Huang, R., Fan, Y. & Lin, Y. Simultaneous detection of multiple cytokines from conditioned media and patient's sera by an antibody-based protein array system. *Anal. Biochem.* **294**, 55–62 (2001).
9. Wiese, R., Belosludtsev, Y., Powdrill, T., Thompson, P. & Hogan, M. Simultaneous multianalyte ELISA performed on a microarray platform. *Clin. Chem.* **47**, 1451–1457 (2001).
10. Jenison, R., La, H., Haeberli, A., Ostroff, R. & Polisky, B. Silicon-based biosensors for rapid detection of protein or nucleic acid targets. *Clin. Chem.* **47**, 1894–1900 (2001).
11. Tam, S.W., Wiese, R., Lee, S., Gilmore, J. & Kumble, K.D. Simultaneous analysis of eight human Th1/Th2 cytokines using microarrays. *J. Immunol. Methods* **261**, 157–165 (2002).
12. Schweitzer, B. *et al.* Multiplexed protein profiling on microarrays by rolling-circle amplification. *Nature Biotechnol.* **20**, 359–365 (2002).
13. Lizardi, P. *et al.* Mutation detection and single-molecule counting using isothermal rolling circle amplification. *Nature Genet.* **19**, 225–232 (1998).
14. Haab, B.B., Dunham, M.J. & Brown, P.O. Protein microarrays for highly parallel detection and quantitation of specific proteins and antibodies in complex solutions. *Genome Biol.* **2**, research0004.1–0004.13 (2001).
15. Sreekumar, A. *et al.* Profiling of cancer cells using protein microarrays: discovery of novel radiation-regulated proteins. *Cancer Res.* **61**, 7585–7593 (2001).
16. Knezevic, V. *et al.* Proteomic profiling of the cancer microenvironment by antibody arrays. *Proteomics* **1**, 1271–1278 (2001).
17. Petricoin, E.F., Zoon, K.C., Kohn, E.C., Barrett, J.C. & Liotta, L.A. Clinical proteomics: translating benchside promise into bedside reality. *Nat. Rev. Drug Discov.* **1**, 683–695 (2002).
18. Belov, L., de la Vega, O., dos Remedios, C.G., Mulligan, S.P. & Christopherson, R.I. Immunophenotyping of leukemias using a cluster of differentiation antibody microarray. *Cancer Res.* **61**, 4483–4489 (2001).
19. Paweletz, C.P. *et al.* Reverse phase protein microarrays which capture disease progression show activation of pro-survival pathways at the cancer invasion front. *Oncogene* **20**, 1981–1989 (2001).
20. Madoz-Gurpide, J., Wang, H., Misek, D.E., Brichory, F. & Hanash, S.M. Protein based microarrays: a tool for probing the proteome of cancer cells and tissues. *Proteomics* **1**, 1279–1287 (2001).
21. Mendoza, L.G. *et al.* High-throughput microarray-based enzyme-linked immunosorbent assay (ELISA). *Biotechniques* **27**, 778–788 (1999).
22. Wiltshire, S. *et al.* Detection of multiple allergen-specific IgEs on microarrays by immunoassay with rolling circle amplification. *Clin. Chem.* **46**, 1990–1993 (2000).
23. Mullenix, M.C., Wiltshire, S., Shao, W., Kitos, G. & Schweitzer, B. Allergen-specific IgE detection on microarrays using rolling circle amplification: correlation with *in vitro* assays for serum IgE. *Clin. Chem.* **47**, 1926–1929 (2001).
24. Kim, T.E. *et al.* Quantitative measurement of serum allergen-specific IgE on protein chip. *Exp. Mol. Med.* **34**, 152–158 (2002).
25. Avseenko, N.V., Morozova, T.Y., Ataullakhanov, F.I. & Morozov, V.N. Immunoassay with multicomponent protein microarrays fabricated by electrospray deposition. *Anal. Chem.* **74**, 927–933 (2002).
26. Hiller, R. *et al.* Microarrayed allergen molecules: diagnostic gatekeepers for allergy treatment. *FASEB J.* **16**, 414–416 (2002).

27. Joos, T.O. *et al.* A microarray enzyme-linked immunosorbent assay for autoimmune diagnostics. *Electrophoresis* **21**, 2641–2650 (2000).
28. Robinson, W.H. *et al.* Autoantigen microarrays for multiplex characterization of autoantibody responses. *Nature Med.* **8**, 295–301 (2002).
29. Mezzasoma, L. *et al.* Antigen microarrays for serodiagnosis of infectious diseases. *Clin. Chem.* **48**, 121–130 (2002).
30. Fromont-Racine, M., Rain, J.C. & Legrain, P. Toward a functional analysis of the yeast genome through exhaustive two-hybrid screens. *Nature Genet.* **16**, 277–282 (1997).
31. Uetz, P. *et al.* A comprehensive analysis of protein–protein interactions in *Saccharomyces cerevisiae*. *Nature* **403**, 623–627 (2000).
32. Ito, T. *et al.* A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl Acad. Sci. USA* **98**, 4569–4574 (2001).
33. Gavin, A.C. *et al.* Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415**, 141–147 (2002).
34. Ho, Y. *et al.* Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* **415**, 180–183 (2002).
35. Ge, H. UPA, a universal protein array system for quantitative detection of protein–protein, protein–DNA, protein–RNA and protein–ligand interactions. *Nucleic Acids Res.* **28**, e3.i–e3.vii (2000).
36. Arenkov, P. *et al.* Protein microchips: use for immunoassay and enzymatic reactions. *Anal. Biochem.* **278**, 123–131 (2000).
37. MacBeath, G. & Schreiber, S.L. Printing proteins as microarrays for high-throughput function determination. *Science* **289**, 1760–1763 (2000).
38. Zhu, H. *et al.* Analysis of yeast protein kinases using protein chips. *Nature Genet.* **26**, 283–289 (2000).
39. Zhu, H. *et al.* Global analysis of protein activities using proteome chips. *Science* **293**, 2101–2105 (2001).
40. Schultz, J., Milpetz, F., Bork, P. & Ponting, C.P. SMART, a simple modular architecture research tool: identification of signaling domains. *Proc. Natl Acad. Sci. USA* **95**, 5857–5864 (1998).
41. Bussow, K. *et al.* A method for global protein expression and antibody screening on high-density filters of an arrayed cDNA library. *Nucleic Acids Res.* **26**, 5007–5008 (1998).
42. Bussow, K., Nordhoff, E., Lubbert, C., Lehrach, H. & Walter, G. A human cDNA library for high-throughput protein expression screening. *Genomics* **65**, 1–8 (2000).
43. Lueking, A. *et al.* Protein microarrays for gene expression and antibody screening. *Anal. Biochem.* **270**, 103–111 (1999).
44. Lee, J. & Bedford, M.T. PABP1 identified as an arginine methyltransferase substrate using high-density protein arrays. *EMBO Rep.* **3**, 268–273 (2002).
45. Walhout, A.J. *et al.* GATEWAY recombinational cloning: application to the cloning of large numbers of open reading frames or ORFeomes. *Methods Enzymol.* **328**, 575–592 (2000).
46. Brizuela, L., Braun, P. & LaBaer, J. FLEXGene repository: from sequenced genomes to gene repositories for high-throughput functional biology and proteomics. *Mol. Biochem. Parasitol.* **118**, 155–165 (2001).
47. Albala, J.S. *et al.* From genes to proteins: high-throughput expression and purification of the human proteome. *J. Cell. Biochem.* **80**, 187–191 (2000).
48. Braun, P. *et al.* Proteome-scale purification of human proteins from bacteria. *Proc. Natl Acad. Sci. USA* **99**, 2654–2659 (2002).
49. Chen, R. *et al.* Simultaneous quantification of six human cytokines in a single sample using microparticle-based flow cytometric technology. *Clin. Chem.* **45**, 1693–1694 (1999).
50. Ziauddin, J. & Sabatini, D.M. Microarrays of cells expressing defined cDNAs. *Nature* **411**, 107–110 (2001).